

Recuperação de informação no ambiente acadêmico: georreferenciamento dos dados dos estudantes do Instituto de Educação Continuada da PUC Minas

Magali Rezende Gouvêa Meireles

***Professora do Departamento de Matemática e
Estatística e Coordenadora Técnico-Pedagógica
da Diretoria de Educação Continuada da Pontifícia
Universidade Católica de Minas Gerais***

Paulo Eduardo Maciel de Almeida

***Professor do Departamento de Computação e
Pesquisador do Laboratório de Sistemas
Inteligentes / CEFET MG***

Ana Carolina Milagres Resende Silva

***Graduada pelo curso de Sistemas de Informação
da Pontifícia Universidade Católica de Minas
Gerais***

A recuperação de informação é a área que estuda a representação e o acesso eficiente à informação. A forma como os dados são apresentados aos usuários pode se tornar um diferencial importante nas rotinas de apoio à decisão. Este artigo apresenta um estudo de caso que utiliza técnicas para o tratamento de dados geográficos, destacando a relevância da representação da informação por meio de mapas temáticos.

Palavras-chave: *Recuperação de Informação; Sistemas de Informação Geográfica; Mapas Temáticos.*

Information retrieval in the academic environment: GIS application of students data at Instituto de Educação Continuada da PUC Minas

Information retrieval concerns representation and efficient access to information. The way data is presented to users is a matter of great importance to decision support routines. This article presents a case study which uses geographical information retrieval techniques and also presents the relevance of information representation through the use of thematic maps.

Keywords: *Information Retrieval; Geographical Information Systems; Thematic Maps.*

Recebido em 23.09.2009 Aceito em 04.12.2009

1 Introdução

Um sistema de informação, automatizado ou manual, que envolve pessoas e máquinas, desenvolve e utiliza métodos de coleta, processamento, transmissão e disseminação de dados. Uma vantagem importante de se utilizar este ambiente integrado é o acesso rápido e seguro, garantindo integridade e estabilidade às informações recuperadas.

Lancaster e Warner (1993), citados por Souza (2006), afirmam que os Sistemas de Recuperação de Informação, SRIs, são a interface entre uma coleção de recursos de informação, em meio impresso ou não, e uma população de usuários, e desempenham as tarefas de aquisição, armazenamento, organização, controle e distribuição de documentos aos usuários.

Os dados processados podem ser apresentados aos usuários em formatos visuais como texto, gráfico ou tabelas. A forma como os dados são apresentados aos usuários tem se tornado critério importante nas rotinas de apoio à decisão. As tecnologias mais conhecidas incluem imagens digitais, Sistemas de Informação Geográfica, interfaces gráficas com o usuário, tabelas e gráficos multidimensionais, realidade virtual, vídeos e animações.

Um Sistema de Informação Geográfica (SIG) é um sistema de computação aplicado para capturar, integrar, manipular e exibir dados usando mapas digitalizados (TURBAN; RAINER; POTTER, 2007). Cada registro possui uma localização geográfica identificada, o que possibilita uma apresentação diferenciada dos dados, facilitando as etapas de planejamento, resolução de problemas e tomadas de decisão.

O objetivo deste artigo é ilustrar a utilização de técnicas matemáticas e computacionais, que relacionam bancos de dados específicos e os apresentam ao usuário no formato de mapas temáticos, tornando mais eficiente a análise das informações recuperadas.

Na seção 2, são apresentados conceitos relacionados ao geoprocessamento e exemplos que utilizam os Sistemas de Informação Geográfica. Na seção 3, avalia-se a utilização de um SIG como ferramenta para a recuperação de informação. São apresentadas, na seção 4, algumas considerações que tornaram o trabalho relevante e passível de ser implementado para a melhoria das condições apresentadas ao usuário.

2 Sistemas de Informação Geográfica

Segundo Câmara, Davis e Monteiro (2001), a coleta de informações sobre a distribuição geográfica de recursos minerais, propriedades, animais e plantas sempre foi uma parte importante das atividades das sociedades organizadas. Até recentemente, no entanto, isto era feito apenas em documentos e mapas em papel, o que impedia uma análise que combinasse mapas e dados. Com o desenvolvimento simultâneo da tecnologia de informática, na segunda metade do século passado tornou-se possível armazenar e representar tais informações em ambiente computacional.

2.1 Traduzindo a informação geográfica para o computador

Para se traduzir a informação geográfica para o computador, Gomes e Velho, citados por Câmara, Davis e Monteiro (2001), utilizam o "paradigma dos quatro universos", que distingue o universo do mundo real, o universo matemático, o universo de representação e o universo de implementação, como apresentado na FIG. 1.

O universo do mundo real inclui as entidades da realidade a serem modeladas no sistema. Encontram-se aqui fenômenos que devem ser representados como tipos de solo, cadastro urbano e rural, dados geofísicos e topográficos.

O universo matemático engloba uma definição matemática formal das entidades a serem representadas. Distinguem-se, neste universo, as classes formais de dados geográficos (dados contínuos e objetos individualizáveis).

O universo de representação é aquele onde as diversas entidades formais são mapeadas para representações geométricas e alfanuméricas no computador. As entidades formais são associadas a diferentes representações geométricas, que podem variar conforme a escala e a projeção cartográfica escolhida, e a época de aquisição do dado. Distinguem-se, aqui, as representações matricial e vetorial.

No universo de implementação, as estruturas de dados e algoritmos são escolhidas baseadas em considerações como desempenho, capacidade

do equipamento e tamanho da massa de dados. O modelo de dados é criado pela utilização das linguagens de programação.

Interface com o usuário

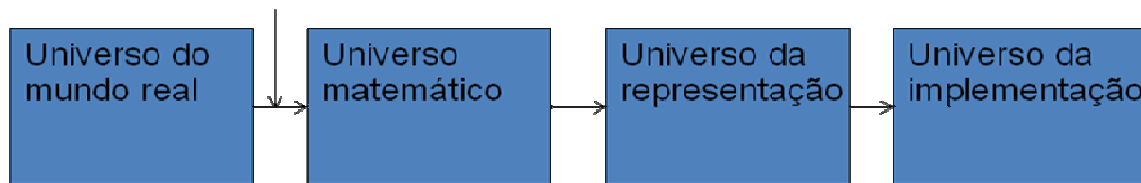


FIGURA 1 - Paradigma dos quatro universos

2.2 Geoprocessamento

O termo Geoprocessamento caracteriza a disciplina do conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento da informação geográfica. As ferramentas utilizadas integram dados de diversas fontes e criam bancos de dados georreferenciados, tornando possível a produção de documentos cartográficos informatizados.

Entende-se por cartografia a ciência que trata da concepção, produção, difusão, utilização e estudo dos mapas, envolvendo estudos e operações científicas, técnicas e artísticas. Um mapa ou carta representa graficamente uma superfície plana, em uma escala definida, destacando suas características naturais ou artificiais.

Para se referenciar um objeto em um mapa, utilizam-se as coordenadas, que são informações lineares ou angulares de um ponto. A escala utilizada na representação define a relação entre o tamanho real da superfície e sua dimensão no mapa. Georreferenciar é atribuir coordenadas a um ponto de um mapa, vinculando-o a um sistema de coordenadas real.

Os dados alfanuméricos são, geralmente, armazenados em tabelas que contêm as informações textuais e numéricas das entidades estudadas. Os nomes dos alunos de uma Instituição e respectivos endereços são exemplos de dados alfanuméricos. Eles podem ser georreferenciados se vinculados às coordenadas terrestres do local a que se referem ou aos dados cartográficos que representam esses locais.

Câmara e Medeiros (1998) citam um exemplo tradicional e verídico de como as análises espaciais típicas de geoprocessamento podem auxiliar a solução de problemas. Em 1854, a cidade de Londres vivia uma grave epidemia de cólera em que mais de 500 pessoas haviam morrido. Nessa época, desconheciam-se o modo de disseminação e contágio da doença, mas o médico John Snow acreditava que podia estar relacionado à água consumida pela população. Como a principal fonte de água da época eram os poços de água, Snow realizou o seguinte procedimento: sobre o mapa das ruas e residências da cidade, marcou cada poço de água com um "x" e cada residência onde havia ocorrido morte por causa da doença com um "ponto". Com os dados espacializados no mapa, Snow observou que muitos casos de cólera ocorriam próximos ao poço da Broad Street, e

decidiu lacrar esse poço. Como consequência, houve uma diminuição dos casos de cólera e Snow constatou que o contágio da cólera estava associado à água consumida pela população. Se os endereços dos poços e os casos de cólera tivessem sido listados em uma tabela, o diagnóstico poderia não ter sido encontrado com a mesma eficiência.

2.3 Tipos de dados em geoprocessamento

De acordo com Câmara, Davis e Monteiro (2001), os dados podem ser classificados em:

Dados temáticos: descrevem a distribuição espacial de uma grandeza geográfica, expressa de forma qualitativa, como os mapas de pedologia e a aptidão agrícola de uma região. Esses dados, obtidos a partir de levantamento de campo, são inseridos no sistema por digitalização ou a partir de classificação de imagens;

Dados cadastrais: cada um de seus elementos é um objeto geográfico, que possui atributos e pode estar associado a várias representações gráficas. Os lotes de uma cidade são elementos do espaço geográfico que possuem atributos (dono, localização, valor venal, IPTU devido, etc.) e que podem ter representações gráficas diferentes em mapas de escalas distintas;

Redes: cada objeto geográfico possui uma localização geográfica exata e está sempre associado a atributos descritivos presentes no banco de dados. Em geoprocessamento, o conceito de rede denota informações associadas a serviços de utilidade pública (água, luz e telefone), redes de drenagem e rodovias;

Modelos numéricos de terreno: esse termo é utilizado para denotar a representação quantitativa de uma grandeza que varia continuamente no espaço. São comumente associados à altimetria e podem, também, ser utilizados para modelar unidades geológicas, como teor de minerais ou propriedades do solo ou subsolo;

Imagens: são obtidas por satélites ou fotografias aéreas e representam formas de captura indireta de informação espacial. Armazenados como matrizes, cada elemento de imagem (denominado "pixel") tem um valor proporcional à energia eletromagnética refletida ou emitida pela área da superfície terrestre correspondente.

2.4 Utilização de software

Um SIG é um sistema de computação aplicado para capturar, integrar, manipular e exibir dados usando mapas digitalizados (TURBAN; RAINIER; POTTER, 2007). Cada registro possui uma localização geográfica identificada, o que possibilita uma apresentação diferenciada dos dados, facilitando as etapas de planejamento, resolução de problemas e tomadas de decisão.

Do ponto de vista da aplicação, utilizar um Sistema de Informação Geográfica implica em escolher as representações computacionais mais

adequadas para capturar a semântica de seu domínio de aplicação. Do ponto de vista da tecnologia, desenvolver um SIG significa oferecer um conjunto de estruturas de dados e algoritmos capazes de representar a grande diversidade de concepções do espaço.

O mercado atual apresenta várias opções de soluções de SIG. Dentre elas, existem os *softwares* livres, que possuem o código fonte aberto, sendo livres para reprodução, modificação e distribuição para a comunidade, e os *softwares* proprietários, que são protegidos por algum tipo de patente. O INPE, Instituto Nacional de Pesquisas Espaciais, estabeleceu um grupo para o desenvolvimento de tecnologia de geoprocessamento e sensoriamento remoto e, a partir de 1991, iniciou o desenvolvimento do SPRING, Sistema de Processamento de Informações Georreferenciadas. Este *software* é proprietário e disponibilizado aos seus usuários deste 1997 (CÂMARA; DAVIS; MONTEIRO, 2001). Vários órgãos, entidades do governo, empresas e universidades fazem uso do SPRING como ferramenta para análises, estudos e levantamentos espaciais.

As técnicas de geoprocessamento mostraram-se eficientes na caracterização do agroecossistema cafeeiro da região de São Sebastião do Paraíso, importante região produtora do sul do estado de Minas Gerais (ALVES *et al.*, 2002). Por meio de operações de tabulação no SPRING, avaliou-se quantitativamente a ocupação da cafeicultura nas unidades ambientais de relevo e solo, mostrando a relação entre café e meio físico.

A Companhia Energética de Minas Gerais (CEMIG) desenvolveu um grande sistema de informações georreferenciadas, denominado Projeto GeoCemig, que, além de disponibilizar uma base de dados corporativa com as principais informações de suas instalações, infra-estrutura, meio ambiente e topologia, integrou não só as bases de dados da Empresa como também as de seus parceiros (MOURA, 1999). O projeto facilitou a utilização de informações georreferenciadas dos sistemas elétrico, de hidrometeorologia, de telecomunicações e de distribuição de gás.

3 SIG como ferramenta para a recuperação de informação

O Instituto de Educação Continuada (IEC) da Pontifícia Universidade Católica de Minas Gerais (PUC Minas) é responsável pela coordenação de cursos de pós-graduação *lato sensu* presenciais. A coordenação técnico-pedagógica é responsável pelo recebimento e análise dos projetos pedagógicos, acompanhamento do desenvolvimento dos cursos e certificação dos alunos. O título de especialista é reconhecido e valorizado pelo mercado, o que torna a qualificação indispensável ao aprimoramento acadêmico profissional. No entanto, um grande número de instituições oferece esta modalidade de curso, tornando-se indispensável a apresentação de informações que auxiliem o aluno em sua escolha pelo curso mais adequado ao seu perfil e às suas demandas. Dentro de uma instituição, são oferecidos cursos em unidades localizadas em endereços distintos, o que aumenta o número de variáveis a serem consideradas

pelo usuário, dificultando a absorção de um grande volume de dados e a transformação de informações em conhecimento aplicado à busca.

Um dos grandes problemas enfrentados pelas instituições é acompanhar, rapidamente, as mudanças geradas pelo avanço da tecnologia e pelas exigências do mundo do trabalho. Um problema clássico é o uso eficiente dos dados e das informações disponíveis nos bancos de dados corporativos das instituições. Os bancos de dados possuem as informações necessárias para responder às consultas, mas as informações recuperadas, muitas vezes, não atendem à expectativa do usuário em obter respostas simples e rápidas em um formato visual de apresentação compreensível e eficiente.

Um Sistema de Gerenciamento de Banco de Dados oferece aos usuários acesso aos dados de uma instituição, minimizando problemas como redundância e incoerência, garantindo a integridade e a segurança dos dados. Parte dos registros de uma instituição de ensino descreve as características dos alunos, denominadas atributos. Os SIG, por sua vez, tratam esses atributos associados a elementos gráficos. Os SIG são formados, basicamente, por pelo menos duas tecnologias computacionais, a cartografia digital, que cria elementos gráficos, e os Sistemas Gerenciadores de Bancos de Dados (SGBD), que permitem a construção e a comparação dos atributos desses elementos. Através da fusão destas tecnologias computacionais, é possível criar não só elementos gráficos para representação do espaço, mas também a junção dos atributos a esses elementos.

Com o objetivo de visualizar os endereços dos alunos matriculados nos cursos de Pós-graduação *lato sensu* do Instituto de Educação Continuada, utilizou-se um SIG para a geração de mapas temáticos. A principal meta era obter as regiões em que se concentrava a maior parte dos alunos de uma determinada área de conhecimento, e definir o percentual dos alunos residentes em outras cidades da região metropolitana de Belo Horizonte e do interior do estado de Minas Gerais. Com a expansão das Unidades do IEC, como mostrado na FIG. 2, tornou-se estratégico obter os locais de concentração de residências dos alunos, para que estas Unidades pudessem ser alocadas em pontos que mais se aproximassem destas residências. O oferecimento de cursos específicos em cada unidade, de acordo com os interesses dos alunos potenciais que residem nas imediações, também pode ser um fator relevante para este estudo.



FIGURA 2 - Distribuição das Unidades do IEC nas regionais de Belo Horizonte

Fonte: SILVA, 2007.

A escolha dos dias em que os cursos são ofertados também sofre influência da localização das residências dos alunos. Se o percentual de alunos da área de educação que moram no interior de Minas Gerais, por exemplo, é grande, os alunos são melhor atendidos se os cursos se concentrarem nos finais de semana.

Para a geração dos mapas propostos, foram utilizados os mapas de Belo Horizonte, separados por regionais com a discriminação dos seus bairros, a base de dados georreferenciada com todos os endereços de Belo Horizonte e um componente espacial, um *software* com ferramentas capazes de gerar os mapas temáticos, neste caso o *SPRING*.

A base de dados disponibilizada pela Prefeitura de Belo Horizonte pode ser definida como a Cartografia Digital do Município, ou seja, o Cadastro Técnico Municipal. A versão utilizada foi atualizada em 2006. Segundo Oliveira (2005), a implantação deste cadastro, pela Prefeitura,

constitui um marco que alterou substancialmente a forma de se obter, manter e divulgar as informações urbanas básicas. Este acervo é mantido pela PRODABEL, cujo nome antigo, Empresa de Processamento de Dados do Município de Belo Horizonte, foi alterado para Empresa de Informática e Informação do Município de Belo Horizonte. O cadastro utilizado foi o Cadastro de Logradouros, que contém registros de todos os logradouros da cidade (ruas, avenidas, praças, etc).

Os dados cadastrais dos alunos da Instituição foram coletados no banco de dados do DATA PUC, Divisão de Tecnologia da PUC Minas. Foram disponibilizados aproximadamente oito mil registros dos alunos matriculados no segundo semestre de 2006, em cursos de pós-graduação *lato sensu*, nas unidades do IEC em Belo Horizonte. Acrescentou-se aos dados fornecidos a classificação da área do curso do aluno.

A concentração dos endereços dos alunos do IEC por regional pode ser observada na FIG. 3. É fácil concluir, após a visualização do mapa, que a maioria dos alunos reside na regional Centro-sul, o que torna o local de residência um fator decisivo na escolha do aluno pelo seu curso. Esta visualização permite-nos uma conclusão imediata, o que, certamente, teria dado um trabalho maior caso os resultados da busca tivessem sido apresentados, por exemplo, por meio de uma tabela. É ilustrada, aqui, uma vantagem clara da utilização dos mapas temáticos em relação a um sistema de recuperação de informação convencional.

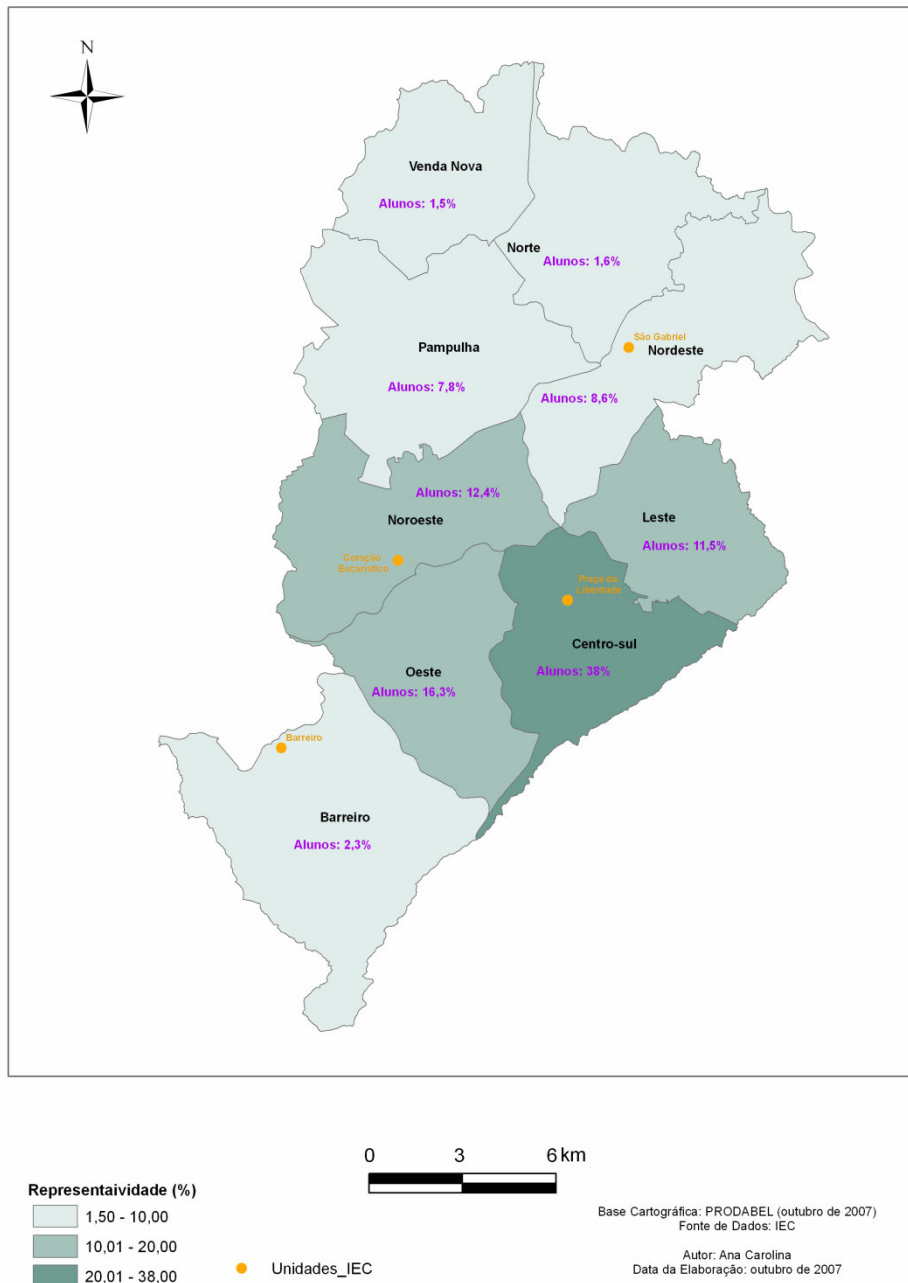


FIGURA 3 - Distribuição das residências dos alunos nas regionais de Belo Horizonte

Fonte: SILVA, 2007.

A partir da observação das áreas de maior concentração nas regionais, como mostrado na FIG. 4, algumas considerações podem ser feitas. A regional Centro-sul, por exemplo, onde localiza-se a sede do IEC, concentra o maior percentual de residências dos alunos da área denominada Direito, em um total de 40,9%.

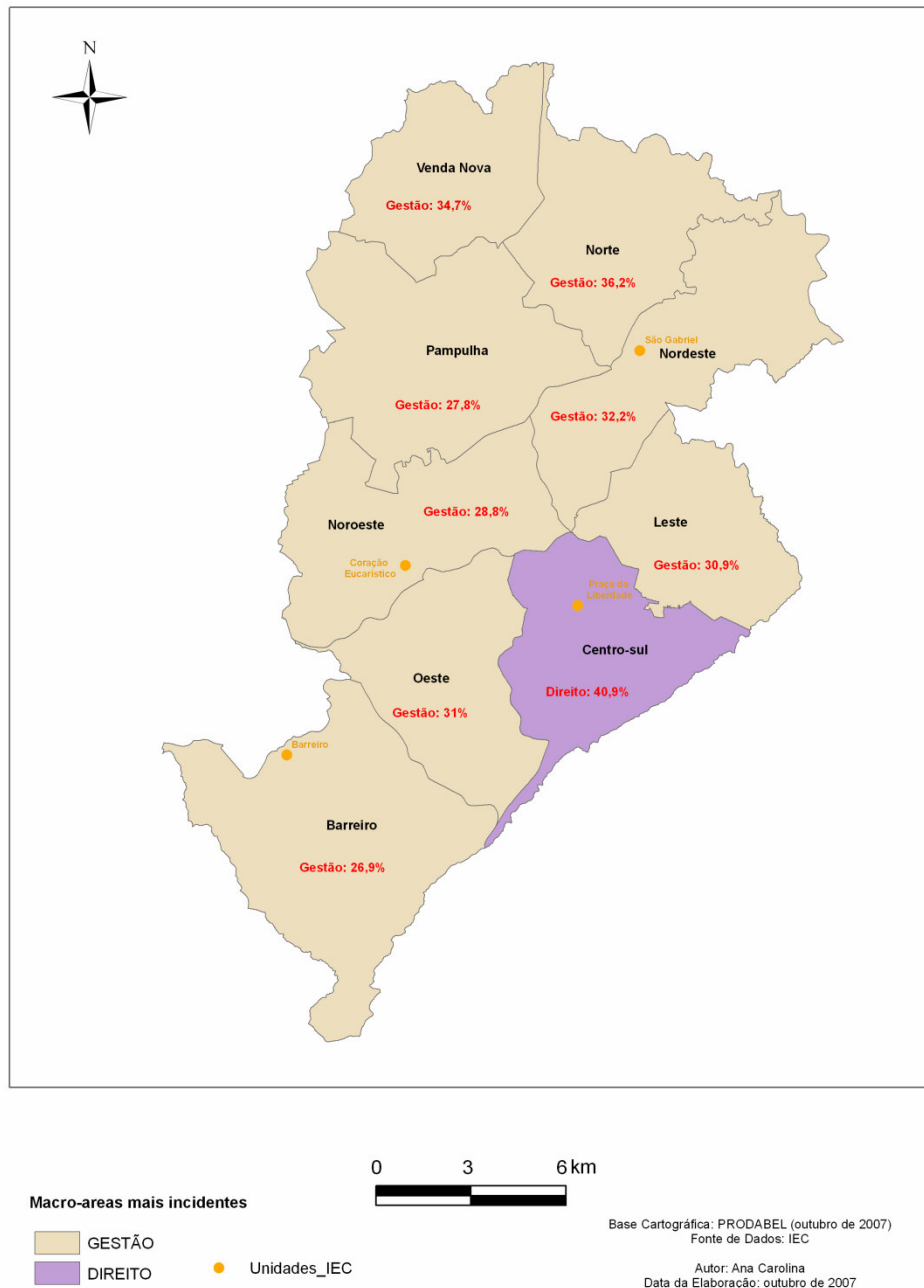


FIGURA 4 - Distribuição das áreas que concentram o maior percentual de residências dos alunos nas regionais

Fonte: SILVA, 2007.

Avaliando-se, separadamente, a regional Centro-sul, podem ser identificados os cursos e as áreas em que concentra-se o maior número de residências dos alunos, como ilustrado pela FIG. 5. Observa-se, assim, que a área predominante é a denominada Direito e o curso que concentra o maior número de alunos residentes nesta regional é o Direito de Empresa.

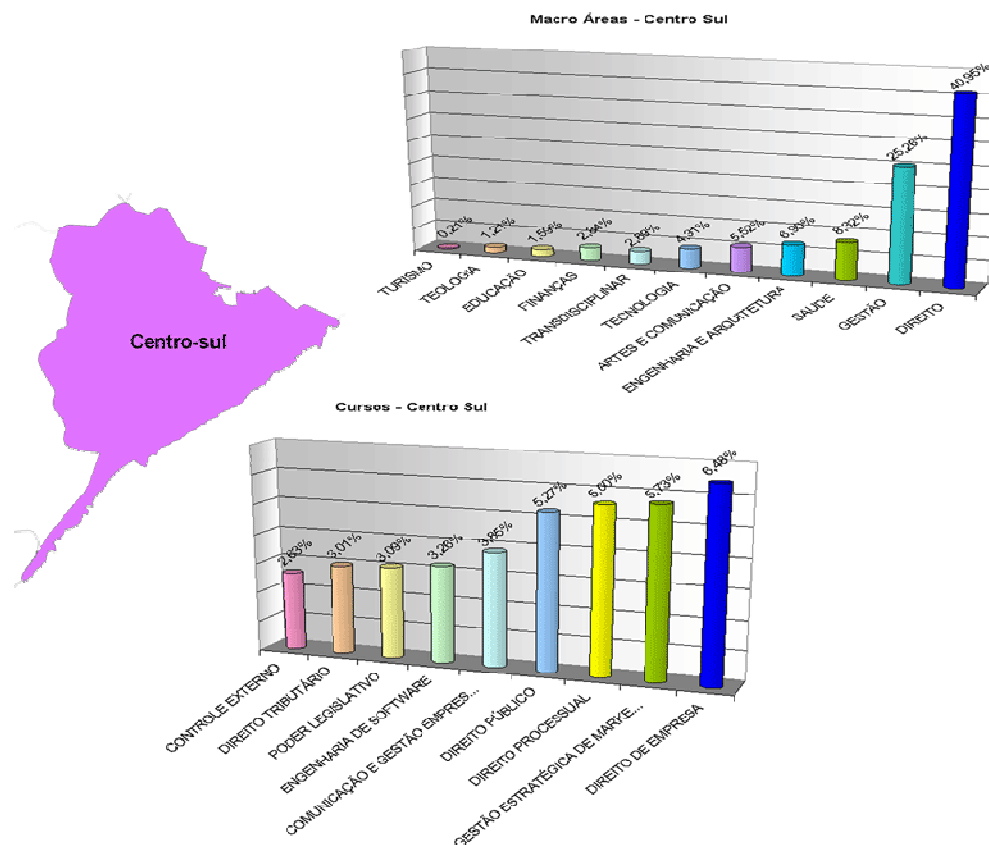


FIGURA 5 - Distribuição da concentração de alunos na regional Centro-sul por área e por curso

Fonte: SILVA, 2007.

4 Considerações finais

Este artigo apresentou uma abordagem diferenciada de emprego de ferramentas de SIG para execução de recuperação de informação em um banco de dados corporativo, com o levantamento de mapas temáticos que auxiliam a análise e a identificação de informações relevantes sobre o cenário atual dos alunos da pós-graduação *lato sensu* presencial da PUC Minas. São informações que passam a ter uma significação diferente quando apresentadas por meio de mapas, e tornam-se, assim, fonte de conhecimento para tomadas de decisão.

Atualmente, inúmeras ferramentas de busca encontram-se disponíveis e, conseqüentemente, variadas formas de apresentação das informações recuperadas são possíveis de ser utilizadas. A apresentação dos dados em um formato alternativo pode conduzir o usuário a análises mais efetivas e determinantes em um processo de apoio à decisão.

Informações como as apresentadas neste trabalho sobre a concentração de residências dos alunos auxiliarão a Instituição, em uma próxima etapa deste trabalho, a definir locais para a instalação de novas Unidades e os melhores dias para se ofertar os cursos. Os dados armazenados em bancos de dados tornam-se compreensíveis se avaliados

em formato gráfico, facilitando, assim, a visualização de informações determinantes para a instituição.

Sem a utilização de SIG e recuperando-se a tabela de endereços dos alunos, o usuário de bancos de dados teria mais dificuldade para agrupar endereços por vizinhança do que, simplesmente, inspecionando pontos georreferenciados plotados em um mapa temático.

Algumas ações, utilizando-se a integração de dados de diversas fontes e bancos de dados georreferenciados, poderão ser implementadas para que a busca do usuário, independentemente de suas habilidades, possa ser considerada satisfatória. No caso do *site* considerado, o endereço do aluno o levará diretamente aos cursos ofertados nas Unidades mais próximas, apresentando, por meio de um mapa, a distância e o melhor caminho entre sua residência e a Instituição.

Referências

ALVES, H. M. R. *et al.* Caracterização de agroecossistemas cafeeiros de Minas Gerais por meio do SPRING. In: SIMPÓSIO BRASILEIRO DE PESQUISAS DOS CAFÉS DO BRASIL, 2., 2002, Vitória. *Anais...* Vitória, 2002. p. 2367-2376.

CÂMARA, N. G.; MEDEIROS, J. S. Princípios básicos em geoprocessamento. In: ASSAD, E. D. *Sistemas de informações geográficas: aplicações na agricultura*. Brasília: Embrapa-SPI/Embrapa-CPAC, 1998. cap. 1, p. 3-11.

CÂMARA, G.; DAVIS, C.; MONTEIRO, A. M. *Introdução a ciência da geoinformação*. [S.l.: S.n.], 2001. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/introd/>. Acesso em: 25 maio 2008.

LANCASTER, F. W.; WARNER, A. J. Information retrieval today. *Information Resources Press*, 1993 *apud* SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em ciência da informação*, Belo Horizonte, v. 11, n. 2, p. 161-173, maio/ago. 2006.

MOURA, C. A. *Otimização dos negócios da CEMIG utilizando geoprocessamento*. 1999. 87 f. Monografia (Pos-graduação *lato sensu* em Geoprocessamento) – Universidade Federal de Minas Gerais, Instituto de Geociências, Belo Horizonte, 1999.

OLIVEIRA, P. A. Usos do cadastro técnico municipal na formulação e execução da política urbana no município de Belo Horizonte. *Revista eletrônica de geografia y ciencias sociales*, Barcelona, v. 9, p. 4-9, 2005.

SILVA, A. C. M. R. *Georreferenciamento dos endereços das residências dos alunos do Instituto de Educação Continuada utilizando o SPRING*. 2007. 52 f. Monografia (Graduação em Sistemas de Informação) – Pontifícia Universidade Católica de Minas Gerais, Sistemas de Informação, Belo Horizonte, 2007.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em ciência da informação*, Belo Horizonte, v. 11, n. 2, p. 161-173, maio/ago. 2006.

TURBAN, E.; RAINER JR., R. K.; POTTER, R. E. *Introdução a sistemas de informação*. Rio de Janeiro: Elsevier, 2007.